8439-EN-01
DTIC

EVALUATION OF GEOSTATISTICS AND WAVELETS FOR IDENTIFYING RELATIONS BETWEEN IMAGERY AND DIFFERENT SPATIAL RESOLUTIONS AND FOR DATA COMPRESSION

Second Interim Report (RSSUSA - **4/2)**

Dr Margaret A. Oliver

January 1999 to May 1999

United States Army

ENVIRONMENTAL RESEARCH OFFICE OF THE U.S. ARMY

London, England

CONTRACT NUMBER - N68171-98-M-5311

Contractor - Approved for Public Release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
01.08.99

3. REPORT TYPE AND DATES COVERED
Interim Jan 99 – May 99

4. TITLE AND SUBTITLE
Evaluation of geostatistics and wavelets for identifying relations between imagery and different spatial resolutions and for data compression

5. FUNDING NUMBERS
N68171
98-M-5311

6. AUTHOR(S)
Dr Margaret A Oliver

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Reading, Whiteknights,
Reading, RG6 2AH, UK

8. PERFORMING ORGANIZATION REPORT NUMBER
RSS USA-4/1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USARDSG-UK, Environmental Sciences Branch
Edison House, 233 Old Marylebone Road,
London, NW1 5TH, UK

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Interim Report:
Summary of work to date

12a. DISTRIBUTION/AVAILABILITY STATEMENT
No limitation on distribution/availability

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This is the second report of the project to apply geostatistics and wavelet analysis to part of the area chosen at For A P Hill in northeastern Virginia. This report embraces a summary of time spent at Reading by J Shine and by Dr Oliver at TEC, cokriging of Korean temperature data, and the analysis of the vegetation data for the first two of four surveys. The cokriging of temperature in Korea is an exercise to determine whether estimates can be improved by using more information on altitude to estimate temperature with smaller error than by ordinary kriging. A detailed statistical and geostatistical analysis has been carried out on the measured vegetation data for surveys 1 and 2. This included a principal components analysis, summary statistics and histograms, and variogram analysis and modelling. Some of the variables showed similar ranges of spatial dependence to those observed in the pixel information from the SPOT image. The cross variograms between some vegetation measures and image data also suggest that there is a relation in their spatial structures.

14. SUBJECT TERMS
Keywords: cokriging, cross variogram. Vegetation, elevation, SPOT

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Attachment 1 Page 1of2

Evaluation of geostatistics and wavelets for identifying relations between imagery and different spatial resolutions and for data compression

Introduction

This report embraces three aspects of recent work: a visit to the University of Reading by James Shine of the Topographic Engineering Center (TEC) for one week in May 1999 and Dr Oliver's visit to TEC in July 1999 for three days; cokriging of Dr P. Krause's temperature data for Korea; and a partial analysis of the ground cover data for A. P. Hill.

Summary of work with James Shine

Dr Oliver and Mr Shine worked together for a week in May 1999 when Mr Shine visited the University of Reading. This time was used for analyses, a draft outline of a proposed paper and discussion. Mr Shine wished go over the analysis for computing the variogram from large sets of data. We experimented with some of the 1-m data for A. P. Hill using the program *ggrid3.f*, written for the project by Professor R. Webster. Mr Shine wanted to develop his experience in this so that he can compute variograms from large data sets within a short time. He left reading feeling confident about this. In addition we also fitted models to the variograms with Genstat and again this reinforced what we did together at TEC last year.

A considerable part of the week was spent discussing the results from the final report of contract No. N68171-97-C-9029 which we now wish to publish. We examined previous issues of the *International Journal of Remote Sensing* to see whether this was suitable for this work. We decided that it was, but that as the content will be small compared with the previous paper we shall submit it as a *Letter*. This is confusing because this form of publication is a short paper in essence and will suit our needs perfectly in this instance. An outline of the paper has been prepared and the introduction written. We shall continue with this when Dr Oliver visits TEC in July.

The remaining time was spent discussing the recent work on the ground survey data. Part of this work is included in this report. However, there is still some way to go on this. We also discussed future work. One idea is to compute a moving variogram to deal with the problems of local trends or non-stationarity in the data. This arises at A. P. Hill for example where there are water bodies and areas of hard standing and buildings. The computer code for this will be written as part of the current contract, but any testing of it will have to be done in the future.

Dr Oliver visited TEC in July 1999 for three days. On arriving she gave a short briefing to Mr W. Clarke (head of section) on the status of our current research, how this builds on work done in the past and where any future research is likely to develop. On the second day Dr Oliver had a meeting with Dr Roper together with Mr Shine. This was to discuss present work and also spatial investigations more generally. Dr Roper invited Dr Oliver to give a general briefing to TEC next year on the research to date.

Part of each day was spent with Mr E. Bosch. We have been exploring a one-dimensional set of radon values in soil where we know there are distinct boundaries. The aim is to see how wavelet analysis deals with this variation and also that of the residuals from the geological classes. We explored different levels of resolution for the raw data. This work is still to be completed.

The work with Mr Shine began by extracting part of the data from the SPOT image and the digital elevation model (DEM). We plan to explore the relations in this smaller file in more detail because statistically the relation between the wavebands and the DEM was weak, yet it was fairly strong for the NIR band visually. The weak relation might arise from the areas of hard standing and buildings which have no particular relation with the elevation. The program ggrid.f would not work with these small files - Mr Shine has since discovered that the zero origin has caused part of the problem.

We continued the discussion about the Letter for IJRS and have decided to use NDVI of subsets from the whole site covered by the 1 m data. This work is being done at present.

Part I

Cokriging temperature data in Korea

The data for the analysis were provided by Dr P. Krause. They comprised temperature and elevation records at 100 sites irregularly scattered over Korea. In addition elevation had been measured at another 565 sites. Table 1 gives the summary statistics for these variables at places where they were both measured. Both have distributions that depart from normality, in particular. Although a geostatistical analysis does not assume that the data are normally distributed it is generally advisable to transform the data to a near-normal distribution for the variogram analysis to stabilize the variances.

Both variables were transformed to common logarithms and for elevation the skewness decreased markedly and the transformed data are close to normal. Temperature departs less so from a normal distribution, but after transformation to common logarithms the departure from normality increases.

Table 1 Summary statistics for Elevation and Temperature

	Elevation	Temperature	Log Elevation	Log Temperature
Number of observations	100	100	100	100
Mean	403.45	53.02	5.17	3.97
Minimum	8.00	33.00	2.08	3.50
Maximum	4546.00	62.00	8.42	4.13
Variance	574928.23	24.95	1.53	0.011
Standard deviation	758.24	4.99	1.24	0.103
Skewness	3.836	-1.45	0.21	-1.93

The data were also examined for trend as part of the exploratory data analysis. This would generally be normal practice when one of the variables is elevation because it can vary in a predictable way. However, in this case it was temperature not elevation whose variation comprised a large element of trend. For elevation linear trend counted for 13.8% of the variation, and quadratic trend for 21.0%. This is much less than expected. It is marginal as to whether this degree of trend should be removed, but it was to ensure that the analysis was reliable. For temperature the trend was much greater: a linear trend accounted for 74.9% of the variation and the quadratic one 77.9%. Clearly a linear trend model is adequate for describing the trend for temperature.

The aim of this analysis was to assess whether temperature could be estimated more reliably with the use of additional information from elevation. In geostatistics the method used is known as cokriging. The value of the method is that it can be used to estimate a property that is more expensive to measure using information from another variable with which it is coregionalized and that is cheaper to measure or that does not change with time. This is particularly true in general for temperature and elevation. There is a physical reason for their relation and elevation does not change substantially in the short term. Therefore, once a digital elevation model has been produced it is a source of inexpensive and reliable information. Cokriging depends on the two (or more) variables being strongly correlated. From the

correlation matrix below it is clear that the correlation between elevation and temperature is moderate.

Table 2. Correlation matrix for temperature and elevation in Korea.

*** Correlation matrix ***

Elevation	1	1.000	
Temperature	2	-0.741	1.000
		1	2

This level of correlation would suggest that it is worthwhile pursuing a coregionalization analysis. The classical correlation coefficient does not take spatial location into account, therefore the relation spatially could be either better or worse.

Cokriging: Theory

The cross variogram

This is the logical extension of ordinary kriging to situations where two or more variables are spatially interdependent or co-regionalized. The first stage is to model the coregionalization. The two regionalized variables, $Z_u(\mathbf{x})$ and $Z_v(\mathbf{x})$, denoted by u and v, both have an autovariogram defined by:

$$\gamma_u(\mathbf{h}) = \frac{1}{2}\mathrm{E}[\{Z_u(\mathbf{x}) - Z_u(\mathbf{x}+\mathbf{h})\}^2]$$

and

$$\gamma_v(\mathbf{h}) = \frac{1}{2}\mathrm{E}[\{Z_v(\mathbf{x}) - Z_v(\mathbf{x}+\mathbf{h})\}^2],$$

and a cross variogram defined as:

$$\gamma_{uv}(\mathbf{h}) = \frac{1}{2}\mathrm{E}[\{Z_u(\mathbf{x}) - Z_u(\mathbf{x}+\mathbf{h})\}\{Z_v(\mathbf{x}) - Z_v(\mathbf{x}+\mathbf{h})\}].$$

The cross variogram function describes the way in which u is related spatially to v. Provided that there are sites where both properties have been measured $\gamma_{uv}(\mathbf{h})$ can be estimated by:

$$\hat{\gamma}_{uv}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})}\sum_{i=1}^{\mathrm{m(h)}}[\{z_u(\mathbf{x}) - z_u(\mathbf{x}+\mathbf{h})\}\{z_v(\mathbf{x}) - z_v(\mathbf{x}+\mathbf{h})\}].$$

which provides the experimental cross variogram for u and v.

The cross variogram can be modelled in the same way as the autovariogram, based on the linear model of coregionalization. Each variable is assumed to be a linear sum of orthogonal random variables $Y(\mathbf{x})$:

$$Z_u(\mathbf{x}) = \sum_{k=1}^{K}\sum_{j=1}^{2} a_{uj}^k Y_j^k(\mathbf{x}) + \mu_{u.}$$

in which

$$\mathrm{E}[Z_u(\mathbf{x})] = \mu_{u.}$$

$$\frac{1}{2}\mathrm{E}[\{Y_j^k(\mathbf{x}) - Y_j^k(\mathbf{x}+\mathbf{h})\}\{Y_{j'}^{k'}(\mathbf{x}) - Y_{j'}^{k'}(\mathbf{x}+\mathbf{h})\}]$$
$$= g_k(\mathbf{h}), \quad \text{positive for } k = k' \text{ and } j = j'$$
$$= 0 \quad \text{otherwise}$$

The variogram for any pair is then:

$$\gamma_{uv}(\mathbf{h}) = \sum_{k=1}^{K}\sum_{j=1}^{2} a_{uj}^k a_{vj}^k g_k(\mathbf{h}).$$

We can replace the products in the second summation by b_{uv}^k to obtain:

$$\gamma_{uv}(\mathbf{h}) = \sum_{k=1}^{K} b_{uv}^k g_k(\mathbf{h}).$$

The variogram for any pair of variables u and v is:

The b_{uv}^k are the nugget and sill variances of the independent components if they are bounded, and for unbounded models they are the nugget variances and gradients.

Cokriging

Once to coregionalization has been modelled it can be used to predict the spatial relations between two or more variables by cokriging. There are generally two reasons for using cokriging:

1. Where one variable is under-sampled compared with another with which it is correlated. The sparsely sampled property can be estimated with greater precision by co-kriging because the spatial information from the more intensely measured one is used in the estimation. The increase in precision depends on the degree of under-sampling and the strength of the coregionalization.

2. When values of all of the variables are known at all sample points, cokriging can improve the coherence between the estimated values by taking account of the relation between them.

If there are V variables, $l = 1,2,\ldots, V$, and the one to be predicted is u, which in our case has been less densely sampled than the others. In ordinary cokriging the estimate is the linear sum:

$$\hat{Z}_u(B) = \sum_{l=1}^{V}\sum_{i=1}^{n_l} \lambda_{il} z_l(\mathbf{x}_i),$$

where the subscript l refers to the variables, of which there are V, and the subscript i refers to the sites, of which there are n_l where the variable l has been measured. The
λ_{il} are the weights, satisfying:

$$\sum_{i=1}^{n_l} \lambda_{il} = 1, \quad l = u; \quad \text{and} \quad \sum_{i=1}^{n_l} \lambda_{il} = 0, \quad l \neq u.$$

These are the non-bias conditions, and subject to them the estimation variance of $\hat{Z}_u(B)$ for a block, B, is minimized by solving equations :

$$\sum_{l=1}^{V}\sum_{i=1}^{n_l} \lambda_{il}\gamma_{vl}(\mathbf{x}_i,\mathbf{x}_j) + \psi_v = \bar{\gamma}_{uv}(\mathbf{x}_j, B) \qquad \text{for all } v\text{=1,2 to } V \text{ and all } j\text{=1,2 to } n_v.$$

The quantity $\gamma_{lv}(x_i, x_j)$ is the cross semivariance between variables l and v at sites i and j, separated by the vector x_i- x_j; $\bar{\gamma}_{uv}(\mathbf{x}_j, B)$ is the average cross semivariance between a site j and the block B, and ψ_v is the Lagrange multiplier for the vth variable. The cokriging variance is obtained from:

$$\sigma_u^2(B) = \sum_{l=1}^{V}\sum_{i=1}^{n_l} \lambda_{jl}\bar{\gamma}_{ul}(\mathbf{x}_j, B) + \psi_u - \bar{\gamma}_{uu}(B,B)$$

where $\bar{\gamma}_{uu}(B,B)$ is the integral of $\gamma_{uu}(\mathbf{h})$ over B, i.e. the within-block variance of u.

Analysis and results of cokriging

Cross variogram

The experimental autovariograms for the raw values of elevation and temperature were computed first. They showed some similarity in their shapes and also ranges of spatial dependence (Figure 1). The autovariograms were then computed on the residuals from the

linear trend for temperature and on the residuals from the quadratic trend for elevation. In addition the elevation was transformed to common logarithms and the variogram was also computed from the transformed data. Considering that the level of skewness is substantial reducing it appears to have had little effect on the variogram. In fact it is less clearly bounded and less related to the variogram of temperature than that for the raw data. The variograms computed from the residuals were more erratic and more difficult to model than those of the raw data. Since the trend appears to be regional in the case of temperature, at the longer lags, I decided to do the analysis on the raw data and the residuals. For kriging it is the first few lags that are important and these are less likely to be affected by the trend than the longer lags.

Although it is important to check the data in this way, the changes did not appear to improve the variogram substantially. This will become evident when the cokriging results are discussed. However, cokriging was carried out on the raw data and the detrended data. During the remaining time on the project I might do some further tests, but I do not expect any major changes.

The experimental auto- and cross-variograms for the raw data are given in Figure1. They have a similar form and the individual autovariograms were fitted best by an exponential model with a distance parameter of about 0.86 units. The same form of model must fit all of the variograms and the range or distance parameter must be the same. The nugget variance, the sills of bounded models and the slope of unbounded models can be different. The coregionalization was modelled by an exponential function with a distance parameter of 0.86 units of latitude and the lower triangle of the sills is given below. The coregionalization of the residuals for elevation and temperature were also modelled and the values used for kriging. The variograms for the residuals were fitted best by a spherical function with a range of 1.01 units of latitude.

a)

b)

c)

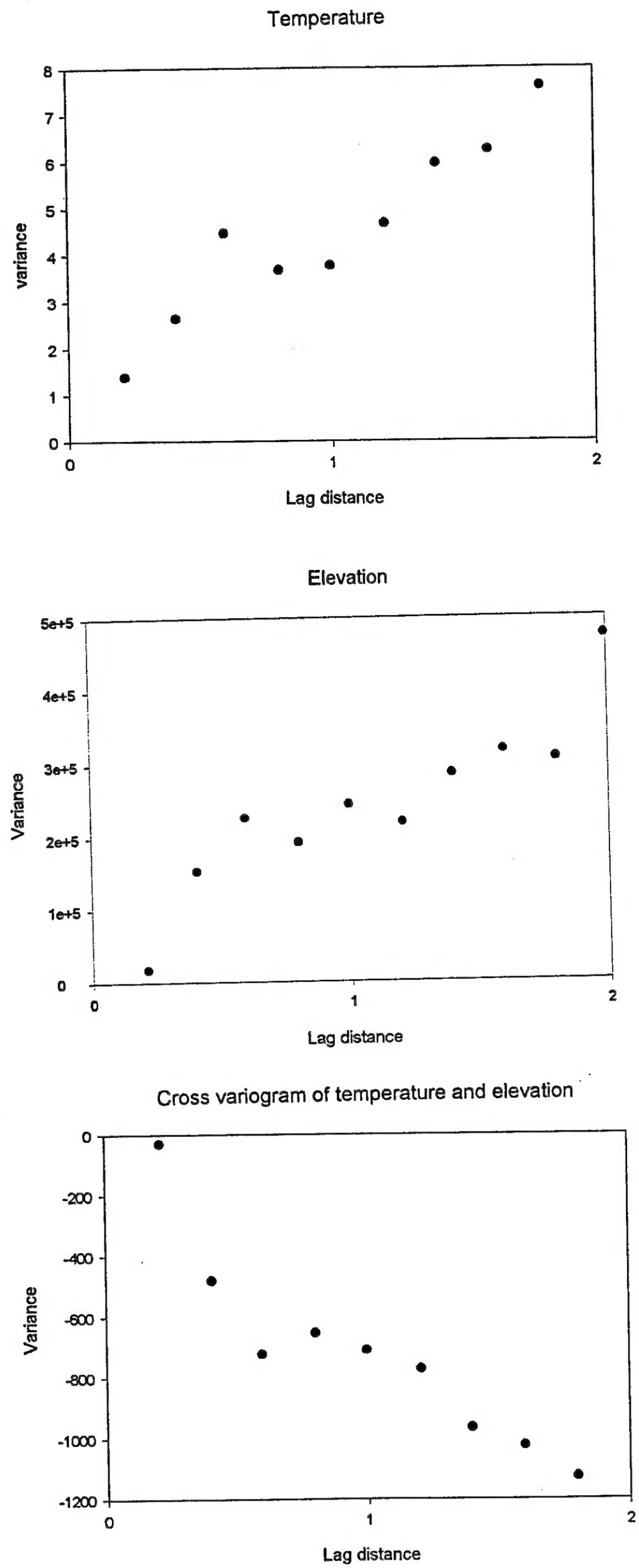


Figure 1: Experimental autovariograms of a) temperature and b) elevation, and c) the experimental cross variogram .

Table 3. Models of coregionalization fitted to the raw data and the residuals from the trend for temperature and elevation.

Fitted sills in lower triangle for the raw data				Fitted sills in lower triangle for the residuals		
Nugget	0.0		Elevation	0.0		Nugget
Variances	0.0	0.6	Cross Temperature	0.0	1.4	variances
Sill	350826.2		Elevation	258385.0		Sill
Variances	-1169.6	6.3	Cross Temperature	-821.4	2.6	variances

Figure 2 shows the experimental cross variograms, the fitted models together with the hull of perfect correlation (the two outer lines). The cross variogram of the residuals coincide with the hull showing a strong correlation. That for the raw data is close to the hull.

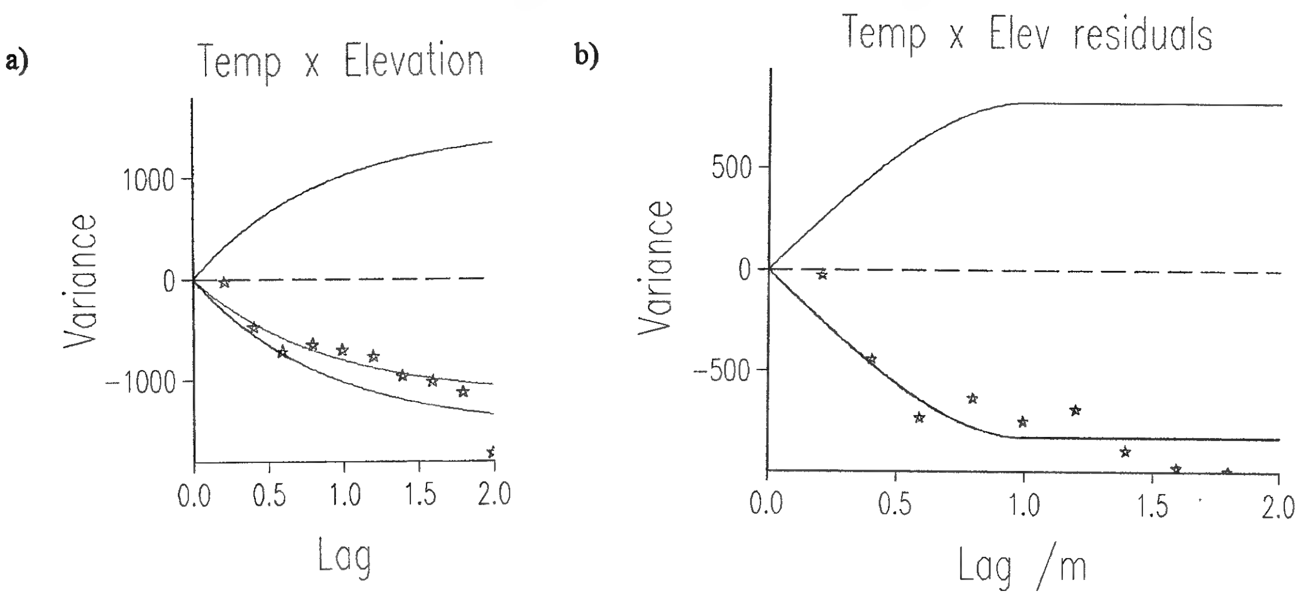


Figure 2: a) Cross variogram of the raw data and b) cross variogram of the residuals, with the hulls of perfect correlation.

Cokriging

The first analysis was to test the modelling and to assess the effects on the estimates of using either the raw data or the residuals. Twenty five of the 100 sites were removed from the raw data and the residuals. Using the models of coregionalization given above the values at the 25 validation points were estimated by punctual cokriging for the raw and residual data, respectively. In addition the raw data were used for autokriging the validation points. The original values, the estimates and the standard errors are given in Table 2.

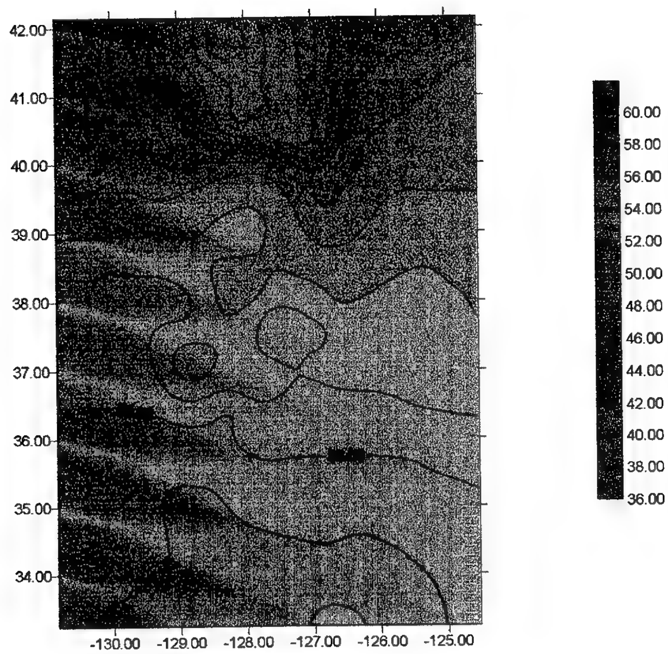
For every validation point the cokriged estimate has a smaller standard error than the autokriged estimate. The differences are small, but they show consistently that cokriging confers a small benefit in terms of estimating temperature more reliably. In addition the estimates are consistently closer to the original values for cokriging of the raw data. For the residuals the standard errors from cokriging are smaller for 15 of the 25 validation points. This was somewhat surprising in relation to the fact that the variograms of the residuals did not appear to be an improvement over that of the raw data. For the residuals the trend was added back so that the values could be compared with the raw data. The estimates are not as consistently good as they are for cokriging with the raw data.

Table 4. Comparison between the raw temperature data, the autokriged estimates and the cokriged estimates, and the cokriged estimates for the residuals and with the trend added back.

		Original	Autokriging		Cokriging		Cokriging residuals		
X	Y	Value	Estimate	SE	Estimate	SE	Estimate	Est+trend	SE
-127.05	37.90	54.0	53.33	2.48	53.32	2.43	0.5392	53.48	2.40
-127.10	37.70	54.0	54.35	1.73	54.25	1.67	0.8142	53.71	1.91
-126.50	33.50	60.0	60.03	1.63	60.07	1.56	-1.3596	58.38	1.82
-128.10	35.20	57.0	57.19	2.19	57.36	2.12	-0.0677	57.96	2.16
-127.75	37.90	53.0	53.86	1.55	53.64	1.49	0.5506	53.74	1.79
-128.00	36.20	54.0	55.17	4.09	55.13	4.08	-0.4151	56.08	3.84
-126.60	37.50	54.0	53.17	2.63	53.24	2.60	0.3207	54.11	2.57
-128.90	37.10	48.0	54.50	4.52	54.47	4.51	-0.0800	55.76	4.03
-129.40	37.00	55.0	55.01	5.43	55.01	5.42	0.1675	56.91	4.31
-126.75	34.30	58.0	58.36	5.29	58.33	5.28	-0.5767	58.26	4.30
-127.65	37.45	56.0	53.82	2.99	53.75	2.97	0.2831	54.34	2.84
-125.65	39.60	50.0	49.56	4.68	49.63	4.66	0.2680	49.52	4.02
-129.01	35.10	59.0	57.95	1.96	57.87	1.93	-0.5659	58.47	2.09
-124.80	40.45	49.0	49.22	4.73	49.36	4.72	0.6924	48.36	3.86
-128.30	41.80	33.0	42.68	4.71	42.71	4.68	-2.4696	40.90	3.72
-128.60	35.90	57.0	56.81	1.71	56.80	1.63	-0.0015	57.47	1.83
-126.50	36.75	54.0	53.60	3.01	53.22	2.97	-1.7979	53.47	2.74
-127.10	37.45	54.0	54.94	1.93	54.89	1.89	1.0006	54.88	2.10
-128.20	36.40	58.0	54.79	3.35	54.76	3.33	-0.4044	55.91	3.15
-127.95	37.40	53.0	53.25	0.98	53.30	0.93	0.1380	54.47	1.53
-129.40	36.03	58.0	56.87	1.50	56.92	1.43	0.5288	58.85	1.69
-124.65	38.00	52.0	51.97	1.66	51.68	1.60	-0.4415	53.86	1.81
-126.40	34.80	58.0	57.45	5.02	57.54	5.00	-0.6245	57.73	3.94
-125.80	39.25	51.0	49.95	4.83	50.00	4.82	0.1461	50.22	4.18
-130.40	42.30	45.0	47.26	6.34	47.38	6.33	-0.3178	45.31	4.23

The entire data set was cokriged as above, but this time using all of the elevation data. The estimates and the standard errors were mapped, Figures 3 to 5. Figures 3a and 4a show the maps of temperature from autokriged and cokriged estimates, respectively. There is remarkably little difference between them. Figure 5a shows the results of cokriging using the residuals and then adding the trend back. This is more different. This appears to show some distortion, however, it is difficult to be certain because we did not have the outline of Korea to superimpose on the estimates. This will be done at TEC. Figures 3b, 4b and 6b show the standard errors for temperature. They are slightly less for cokriging. These values show the pattern of sampling and also the coastline of the country.

Ordinary kriged estimates of temperature for Korea



Standard errors from ordinary kriging of temperature for Korea

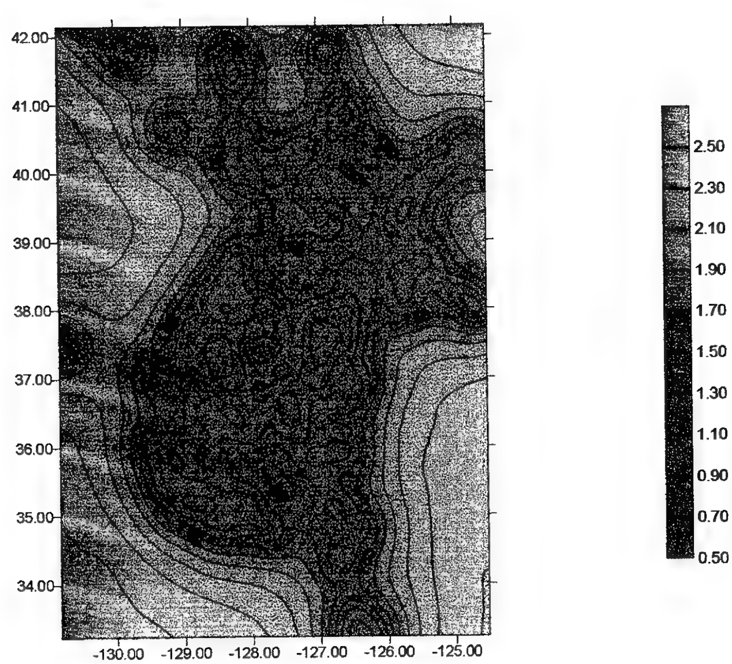


Figure 3: a) Map of estimates from autokriging of temperature for Korea,
b) map of the standard errors from autokriging of temperature

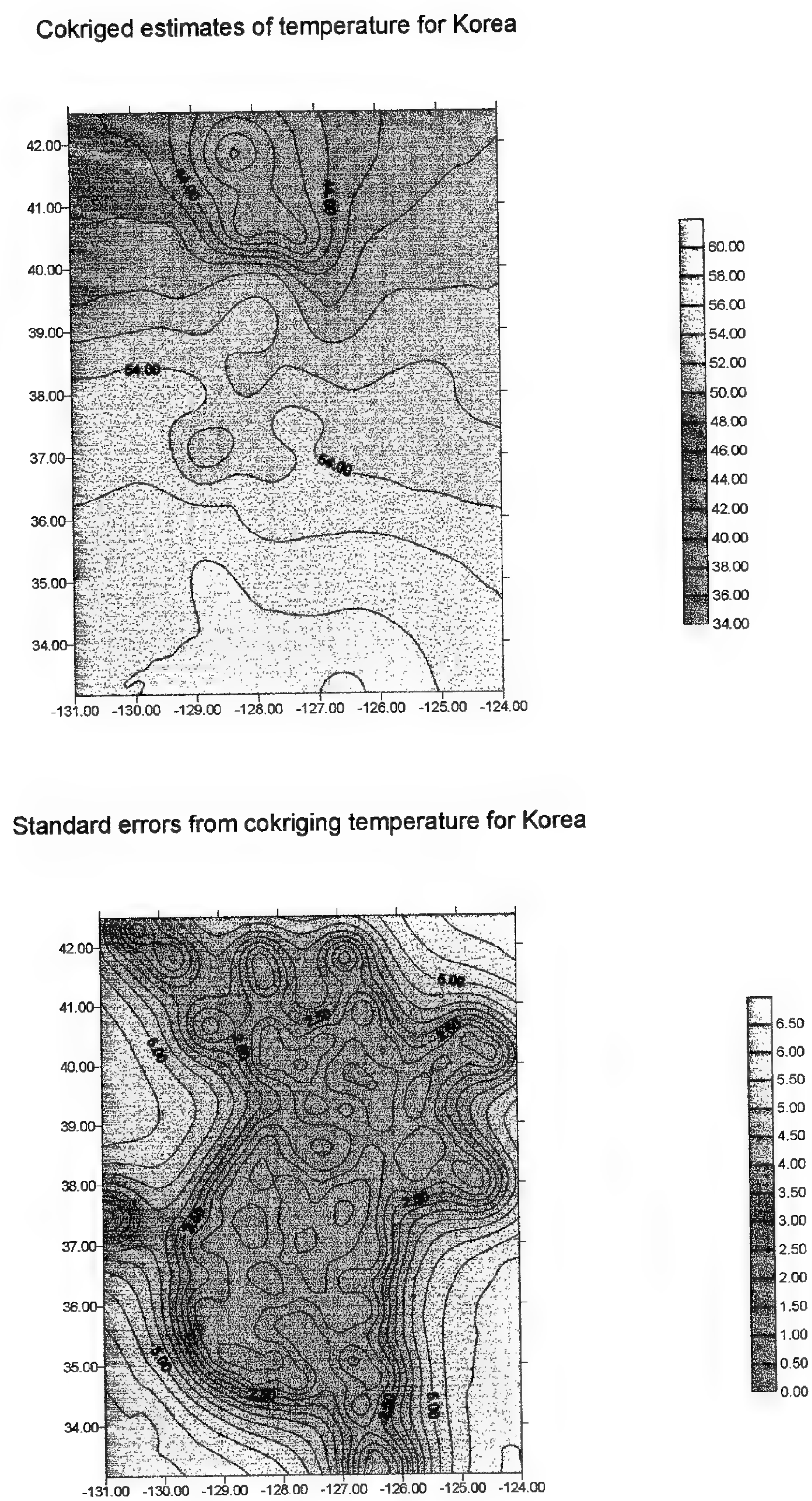
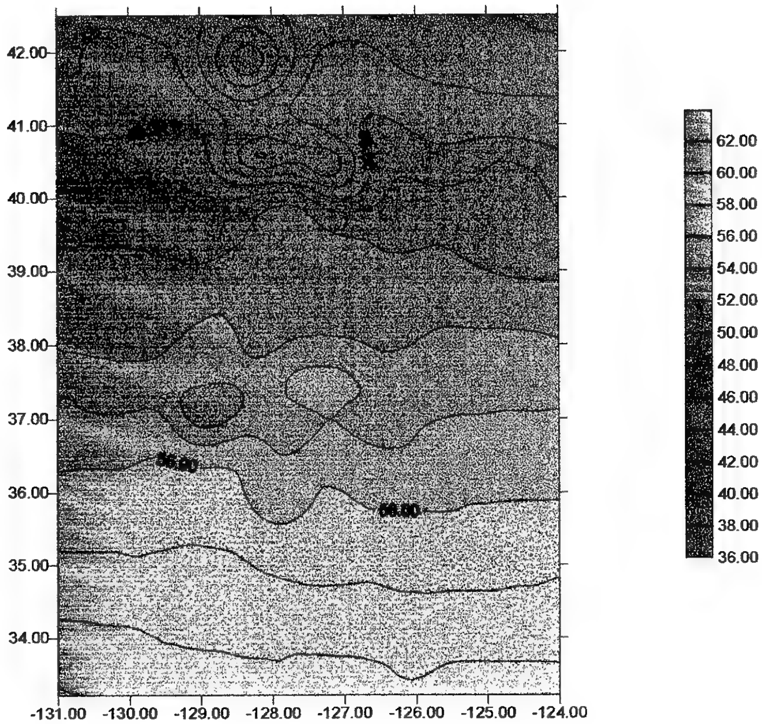


Figure 4: a) Map of cokriged estimates of temperature for Korea,
b) map of the standard errors from cokriging of temperature.

Cokriged estimates of residuals of temperature with trend added back for Korea



Standard errors of residuals from cokriging temperature for Korea

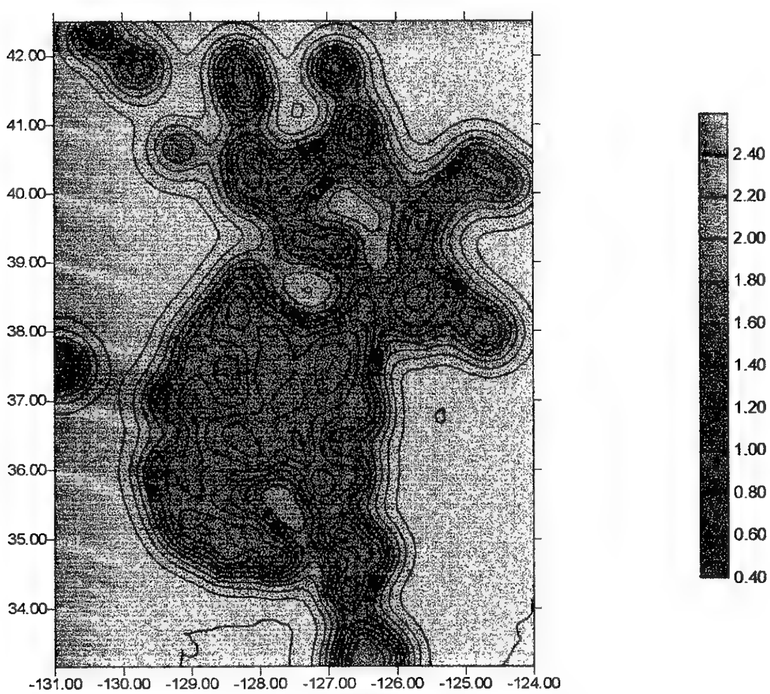


Figure 5: a) Map of the cokriged estimates of the residuals for temperature with trend added back for Korea, b) map of the standard errors from cokriging the residuals of temperature.

Part II

Introduction

In this report the first part of the vegetation analysis will be described. It covers the analysis of the quantitative data in surveys 1 and 2 which have been combined for this analysis in parts. The remaining analyses of the class data for surveys 1, 2 and 4 will be part of the final report.

Surveys 1 and 2

Survey 1 was carried out in 1997 at A. P. Hill. The sample comprises several small transects that have a random starting positions within the seven strata of the training areas. The plot size corresponded with the SPOT pixel size of 20 m by 20 m. The points along the transects were at 100 m intervals (see Figure 6). This survey mainly embraced either hard or soft woodland areas of vegetation. The second survey was a square grid with an interval of 300 m covering the whole of our study site at A. P. Hill (Figure 7). Since there were many sites without quantitative woodland information, because it included grassland, buildings and hard standing, the sites with quantitative information were analysed with the data from the first survey.

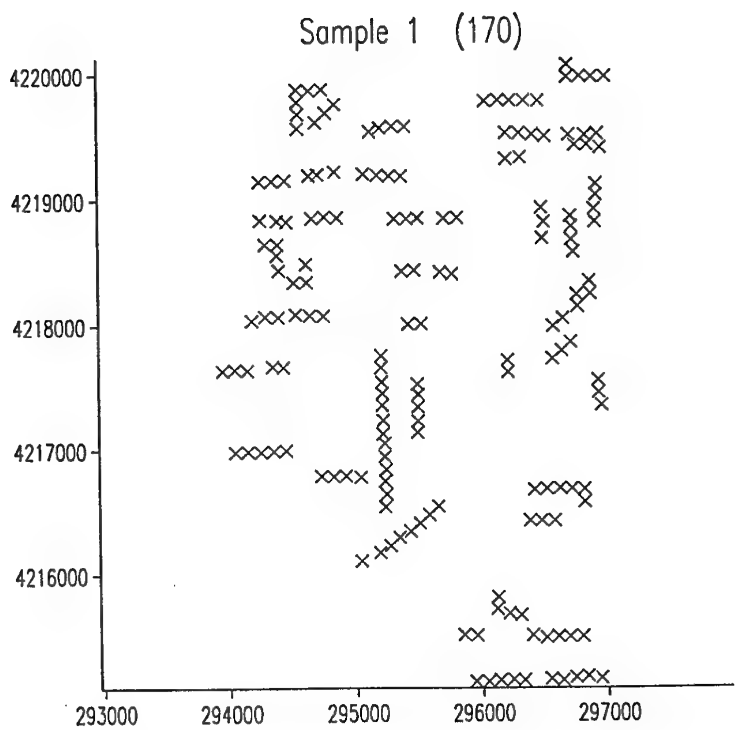


Figure 6: Map of sites for Survey 1.

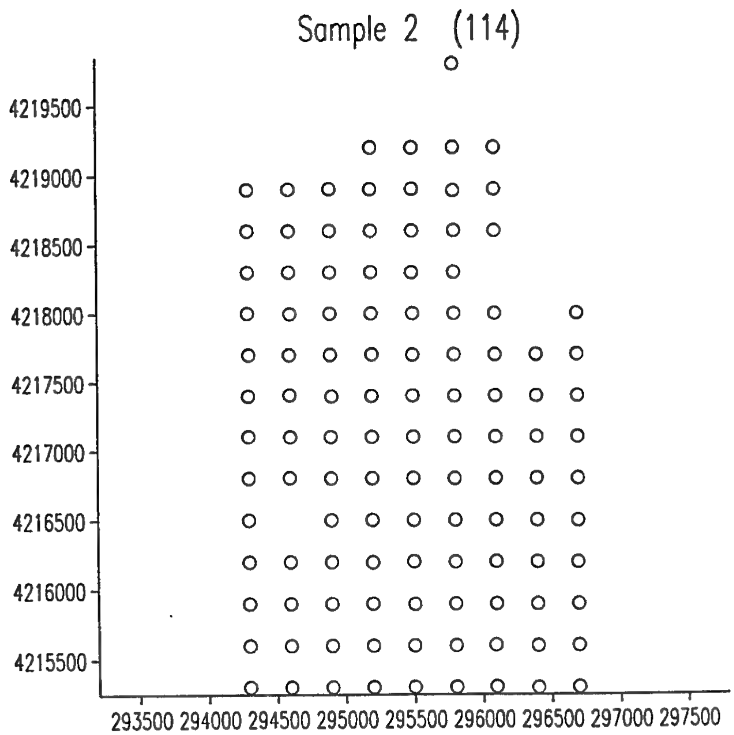


Figure 7: Map of sites for Survey 2.

Exploratory data analysis

The summary statistics of the 17 quantitative variables were analysed for surveys 1 and 2 separately. They are given in Tables 5 and 6. The skewness values are generally small showing that the statistical distribution does not depart seriously from normal, except for stem spacing (survey 1). This variable had one extreme value which was removed to obtain a near-normal distribution for the variogram analysis. Figures 8 and 9 show the histograms of the variables listed below for survey 1. The digital numbers for the three wavebands of the SPOT image that coincided with sites where the vegetation had been examined were also extracted and their summary statistics are given in Table 7 for both surveys. Their histograms are shown in Figure 10.

Variables analysed and their abbreviation:

This part of the list contains those variables related to forest density (Set A):

maxcc - maximum range of visual estimate of crown closure (%)
ovstmin - minimum range of overstory height (ft)
ovstmax - mamimum range of overstory height (ft)
undstmn - minimum range of understory height (ft)
undstmx - maximum range of understory height (ft)

ba_f - estimate of basal area per hectare (metric units)
stem - total stems in plot (count)
ba_tot - sum of all basal area for each tree per plot (square metres)
stemsp - average minimum distance between stems within each plot (metres)

This part of the list contains those variables related to tree species (Set B):

ba_so - percentage of total basal area that is softwood in each plot
ba_ha - percentage of total basal area that is hardwood in each plot
stem_so - percentage of total number of stems that are softwood in each plot
stem_ha - percentage of total number of stems that are hardwood in each plot
bad_so - percentage of dominant basal area that is softwood in each plot
bad_ha - percentage of dominant basal area that is hardwood in each plot
stemd_so - percentage of dominant number of stems that are softwood in each plot
stemd_ha - percentage of dominant number of stems that are hardwood in each plot

Table 5: Summary statistics for vegetation measures for Survey 1

Variable	*N*	*Missing*	*Mean*	*Median*	*Min*	*Max*	*Variance*	*Standard deviation*	*Skewness*	*Kurtosis*
maxcc	169	1	67.04	70.0	0.0	100.0	543.3	23.31	-1.22	0.43
minovst	169	1	73.72	80.0	15.0	110.0	395.1	19.88	-1.19	1.18
maxovst	169	1	78.54	80.0	20.0	110.0	405.1	20.13	-1.32	1.29
minunst	169	1	11.35	10.0	0.0	30.0	33.3	5.77	0.96	2.31
maxunst	169	1	20.66	20.0	0.0	35.0	62.2	7.89	-0.70	0.28
ba_f	168	1	34.36	34.3	2.4	76.2	199.4	14.12	0.05	0.31
stem	169	1	20.44	19.0	0.0	81.0	112.9	10.62	1.96	7.01
bat_tot	169	1	1.07	1.1	0.0	2.4	0.2	0.45	0.01	0.31
stemsp	168	2	2.33	2.2	0.9	7.3	0.6	0.78	2.00	9.10
ba_so	168	2	46.16	43.5	0.0	100.0	1487.1	38.56	0.13	-1.58
ba_ha	168	2	53.54	54.5	0.0	100.0	1487.1	38.56	0.13	-1.58
stem_so	138	2	41.01	33.3	0.0	100.0	1338.9	36.59	0.33	-1.41
stem_ha	168	2	58.99	66.7	0.0	100.0	1338.9	36.59	0.33	-1.41
bad_so	168	2	48.94	46.7	0.0	100.0	1610.3	40.13	0.06	-1.64
bad_ha	168	2	51.06	54.3	0.0	100.0	1610.3	40.13	0.06	-1.64
Stemd_so	168	2	49.14	48.8	0.0	100.0	1570.7	39.63	0.02	-1.62
Stemd_ha	168	2	50.86	51.2	0.0	100.0	1570.7	39.63	0.02	-1.62

Table 6: Summary statistics for vegetation measures for Survey 2

Variable	*N*	*Missing*	*Mean*	*Median*	*Min*	*Max*	*Variance*	*Standard deviation*	*Skewness*	*Kurtosis*
maxcc	60	54	68.17	80.0	5.0	100.0	674.5	25.97	-0.99	-0.02
minovst	0	114	*	*	*	*	*	*	*	*
maxovst	60	54	73.75	80.0	20.0	100.0	315.8	17.77	-1.32	1.52
minunst	17	97	8.59	10.0	1.0	20.0	34.9	5.91	0.28	-1.07
maxunst	54	60	15.11	15.0	3.0	25.0	23.9	4.89	-0.07	-0.31
ba_f	58	56	32.06	34.5	3.4	57.9	175.9	13.26	-0.35	-0.73
stem	58	56	19.91	18.5	5.0	44.0	96.1	9.80	0.63	-0.29
bat_tot	58	56	1.07	1.1	0.1	1.8	0.17	0.42	-0.35	-0.73
stemsp	58	56	2.50	2.4	1.2	5.0	0.57	0.76	1.05	1.19
ba_so	58	56	57.54	63.5	0.0	100.0	1360.3	36.88	-0.31	-1.46
ba_ha	58	56	42.46	36.5	0.0	100.0	1360.3	36.88	-0.31	-1.46
stem_so	58	56	50.33	52.1	0.0	100.0	1288.4	35.89	-0.05	-1.53
stem_ha	58	56	49.67	47.9	0.0	100.0	1288.4	35.89	-0.05	-1.53
bad_so	58	56	60.87	64.9	0.0	100.0	1472.5	38.37	-0.38	-1.43
bad_ha	58	56	39.14	35.1	0.0	100.0	1472.5	38.37	-0.38	-1.43
Stemd_so	58	56	60.15	71.8	0.0	100.0	1520.1	38.99	-0.34	-1.52
Stemd_ha	58	56	39.93	28.2	0.0	100.0	1520.1	38.99	-0.34	-1.52

Table 7: Summary statistics for the three wavebands from SPOT for Surveys 1 and 2

Variable	*N*	*Missing*	*Mean*	*Median*	*Min*	*Max*	*Variance*	*Standard deviation*	*Skewness*	*Kurtosis*
Red (1)	116	54	61.94	61.0	58.0	80.0	14.5	3.81	2.49	7.18
Green (2)	116	54	36.33	34.0	32.0	67.0	33.56	5.79	3.17	11.71
NIR (3)	116	54	119.3	121.0	62..0	148.0	249.5	15.79	-0.55	0.57

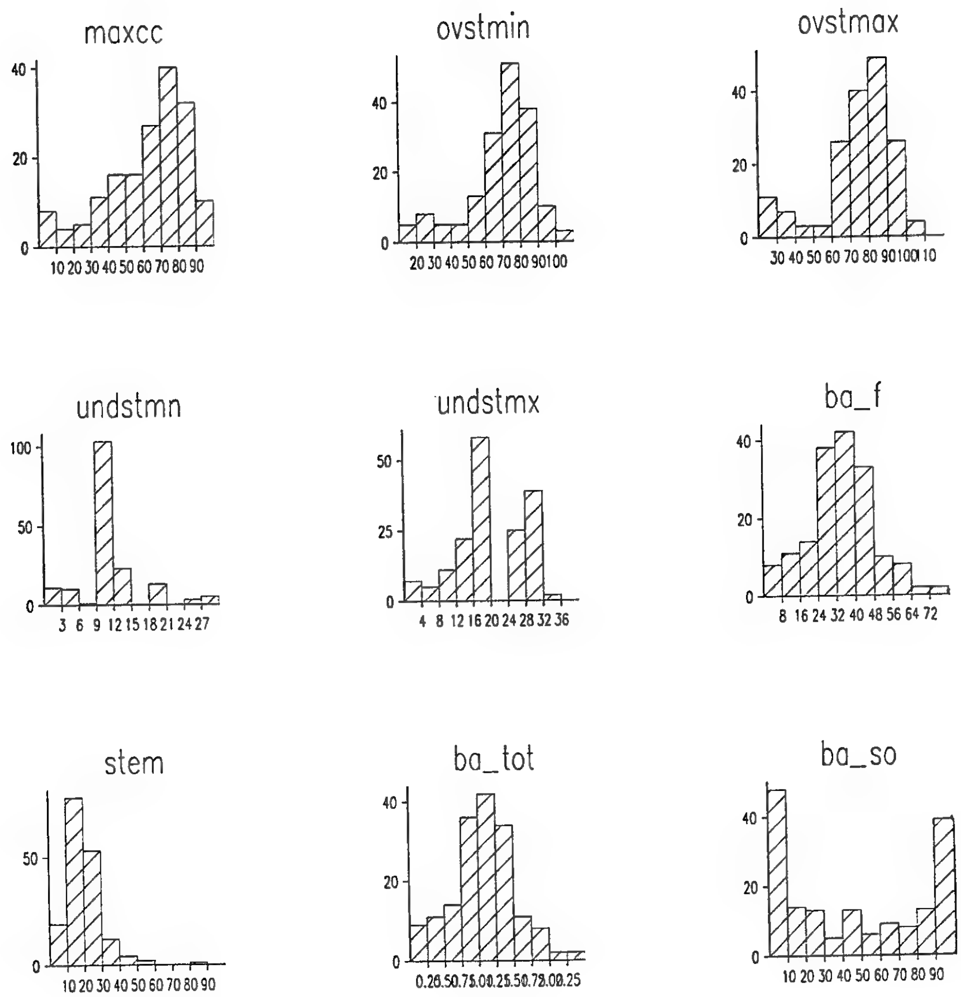


Figure 8: Histograms of variables in Set A of Survey 1.

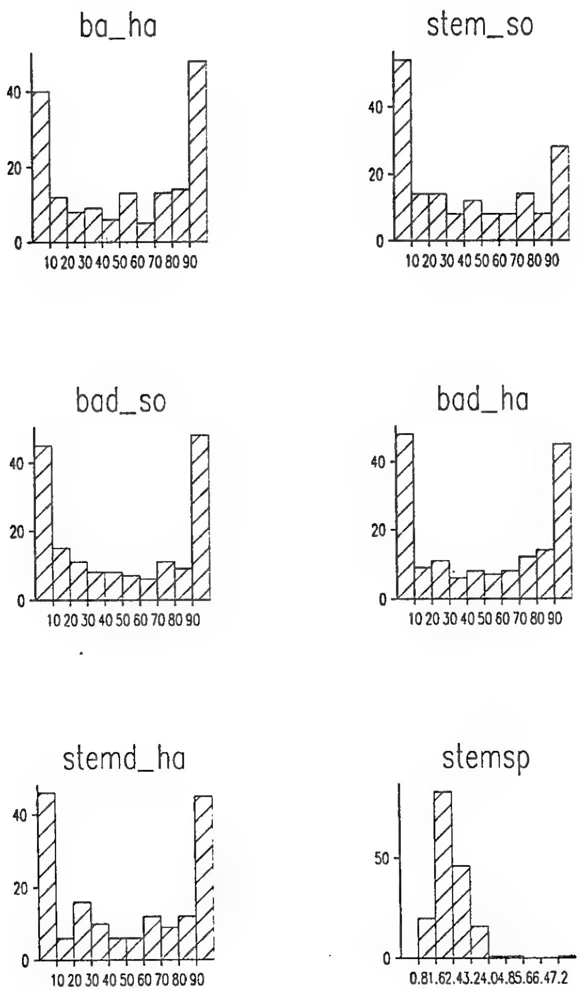


Figure 9: Histograms of variables in Set B of Survey 1.

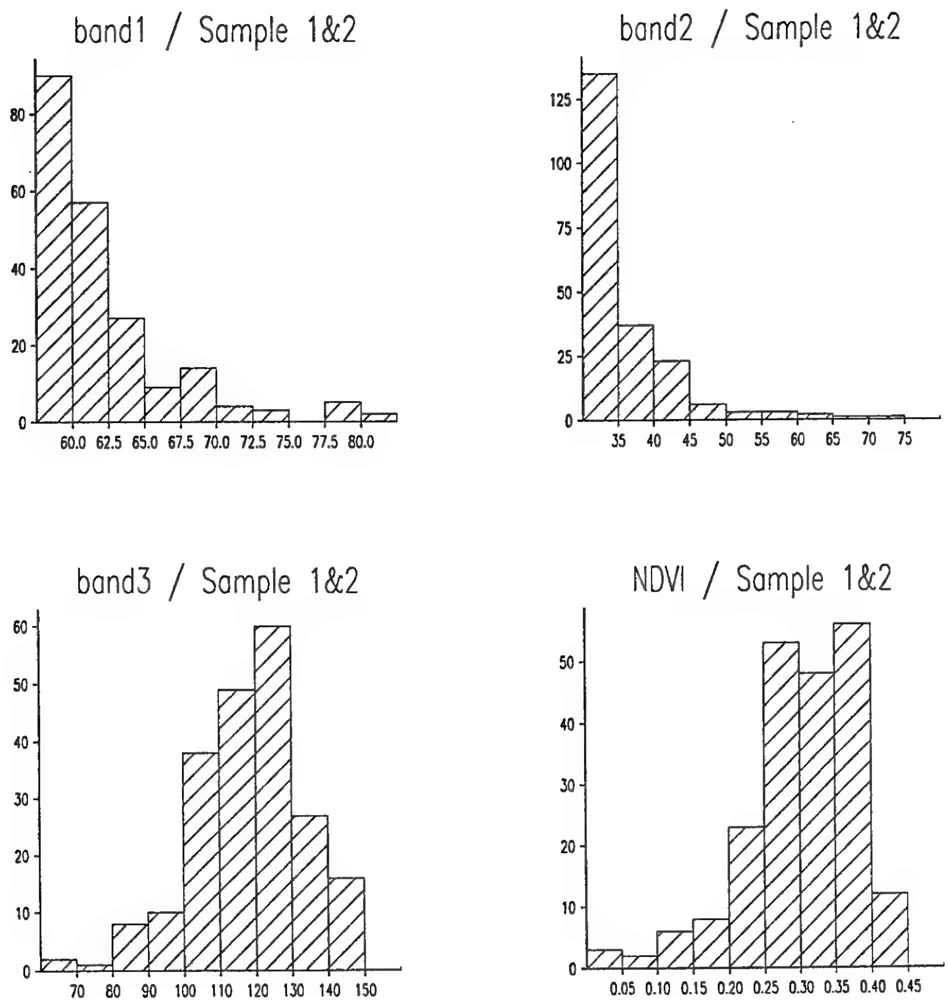


Figure 10: Histograms of wavebands 1 (Red), 2 (Green), 3 (NIR) and NDVI for sites coinciding with Surveys 1 and 2.

To assess which of these variables were likely to represent the variation the data most strongly a principal components analysis was done on the correlation matrix. The latter was used because it effectively standardizes the data. The first component accounted for 53.7% of the variation and the second 18%. The variables that 'loaded' most heavily on the first component were:

ba_so, ba_ha, stem_so, stem_so, stem_ha, bad_so, bad_ha, stemd_so and stem_ha.

The variables that 'loaded' most heavily on the second component were:

maxcc, ba_f, stem, ba_tot and stemsp.

A set of variables that is considered to express the variation and summarise it adequately is:

maxcc, ba_f, stem, stemsp, ovstmax, undstmx and ba_so.

These are based on the distribution of the variables in the plane of PC1 and PC2 (Figure 11).

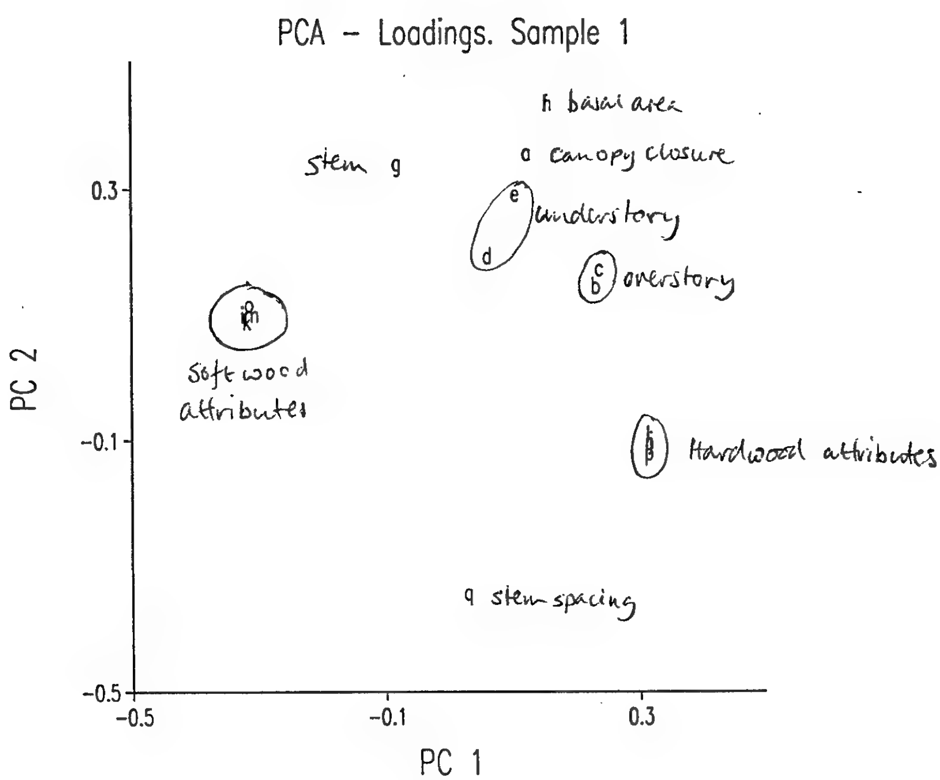


Figure 11: Plot of variables based on their loadings in the plane of PC1 and PC2.

Table 8 gives the correlations for the vegetation measures and the DNs of the wavebands. In general these are small for the vegetation measures and DNs. Those for NIR are the largest for maxcc, stem_so, ste_ha, stemd_so and stem_ha. There are some strong correlations for the vegetation measure which are to be expected, for example ba_so and ba_ha which add to 100%.

Table 8. Correlations for the vegetation measures and the three wavebands from the SPOT image.

*** Correlation matrix ***

	band1	band2	band3	cc	ovstmin	ovstmax	undstmn
band1	1.000						
band2	0.960	1.000					
band3	-0.205	-0.308	1.000				
cc	0.023	0.017	0.227	1.000			
ovstmin	0.036	0.018	0.178	0.205	1.000		
ovstmax	0.030	0.021	0.198	0.237	0.956	1.000	
undstmn	-0.148	-0.113	0.076	0.145	0:196	0.217	1.000
undstmx	0.073	0.080	0.099	0.260	0.427	0.487	0.515
ba_f	0.179	0.177	0.087	0.518	0.557	0.594	0.141
stem	0.081	0.098	-0.021	0.474	-0.346	-0.306	0.021
ba_tot	0.178	0.177	0.087	0.518	0.557	0.594	0.141
ba_so	0.005	-0.005	-0.182	-0.260	-0.596	-0.625	-0.014
ba_ha	-0.005	0.005	0.182	0.260	0.596	0.625	0.014
stem_so	0.004	0.010	-0.224	-0.230	-0.676	-0.687	-0.053
stem_ha	-0.004	-0.010	0.224	0.230	0.676	0.687	0.053
bad_so	0.002	-0.015	-0.158	-0.276	-0.541	-0.572	0.007
bad_ha	-0.002	0.015	0.158	0.276	0.541	0.572	-0.007
stemd_so	0.000	-0.006	-0.211	-0.274	-0.547	-0.568	0.009
stemd_ha	0.000	0.006	0.211	0.274	0.547	0.568	-0.009
stemsp	-0.007	-0.033	-0.030	-0.456	0.182	0.136	-0.015

	undstmx	ba_f	stem	ba_tot	ba_so	ba_ha	stem_so
undstmx	1.000						
ba_f	0.396	1.000					
stem	-0.014	0.260	1.000				
ba_tot	0.396	1.000	0.260	1.000			
ba_so	-0.146	-0.385	0.267	-0.385	1.000		
ba_ha	0.146	0.385	-0.267	0.385	-1.000	1.000	
stem_so	-0.207	-0.385	0.309	-0.385	0.941	-0.941	1.000
stem_ha	0.207	0.385	-0.309	0.385	-0.941	0.941	-1.000
bad_so	-0.116	-0.376	0.243	-0.376	0.990	-0.990	0.894
bad_ha	0.116	0.376	-0.243	0.376	-0.990	0.990	-0.894
stemd_so	-0.121	-0.349	0.269	-0.349	0.972	-0.972	0.939
stemd_ha	0.121	0.349	-0.269	0.349	-0.972	0.972	-0.939
stemsp	-0.174	-0.302	-0.660	-0.302	-0.122	0.122	-0.189

	stem_ha	bad_so	bad_ha	stemd_so	stemd_ha	stemsp
stem_ha	1.000					
bad_so	-0.894	1.000				
bad_ha	0.894	-1.000	1.000			
stemd_so	-0.939	0.966	-0.966	1.000		
stemd_ha	0.939	-0.966	0.966	-1.000	1.000	
stemsp	0.189	-0.076	0.076	-0.137	0.137	1.000

Variogram analysis

Experimental variograms were computed for all of the variables listed above for the combined data from surveys 1 and 2. Variograms were computed in four directions at the outset, but the number of sites is marginal for this. For set A variables the directions of maximum and minimum variation are not consistent, but for set B variables the variation in direction NNE to SSW (o) have the longest range of spatial dependence and the largest sill variances and those at right angles have the shortest ranges and the smaller sill varinaces (*) (Figures 12 and 13).

Figures 14 and 15 show the experimental omnidirectional variograms for the two sets of variables from surveys 1 and 2. Those that show reasonable spatial structure are: maxcc, ovstmin, ovstmax, undstmn, stem, ba_so, ba_ha, stem_so, stem_ha, bad_so, bad_ha, stemd_so and stemd_ha. For the twin variables, such as ba_so and ba_ha the variograms are identical for the reasons given earlier. The following variables were modelled: maxcc, overstory height (derived from ovstmin and ovstmax), understory height (derived from undstmn and undstmx), ba_f, stem, stem spacing, ba_so (equivalent to ba_ha also), stem_so, bad_so and stemd_so. In addition the multivariate variogram from this analysis was computed and modelled, also elevation, and the three wavebands and NDVI. They are shown in Figures 16 to 19.

Table 7 gives the model parameters of the variables modelled. The experimental variograms of many of the properties in Table 7 are somewhat erratic. This could be related to the irregular sampling scheme. However, there appears to be some evidence of periodicity in several variograms with wavelengths of between 500 m and 700 m. A previous report that contained transects of the pixels to match the vegetation ones also showed periodicity in the DNs. There appears to be some relation between the range of spatial dependence of elevation and several of the vegetation measures. The multivariate variogram has identified a short range component of variation of just over 300 m which matches with the short range component of NIR. The variograms of the vegetation classes will be examined in the next report. The models fitted to directional variograms of ba_so are revealing: the variation in direction 135° is 462 m and that in direction 45° is 1271 m. This suggests that the different ranges might reflect some anisotropy in the variation. This was identified in the image data, but because the sill heights were different this signalled zonal anisotropy which cannot be corrected simly. It suggests that there are distinct strata present and this is evident from the areas with different kinds of vegetation. There are also distinct landscape units which will be explored in the next report.

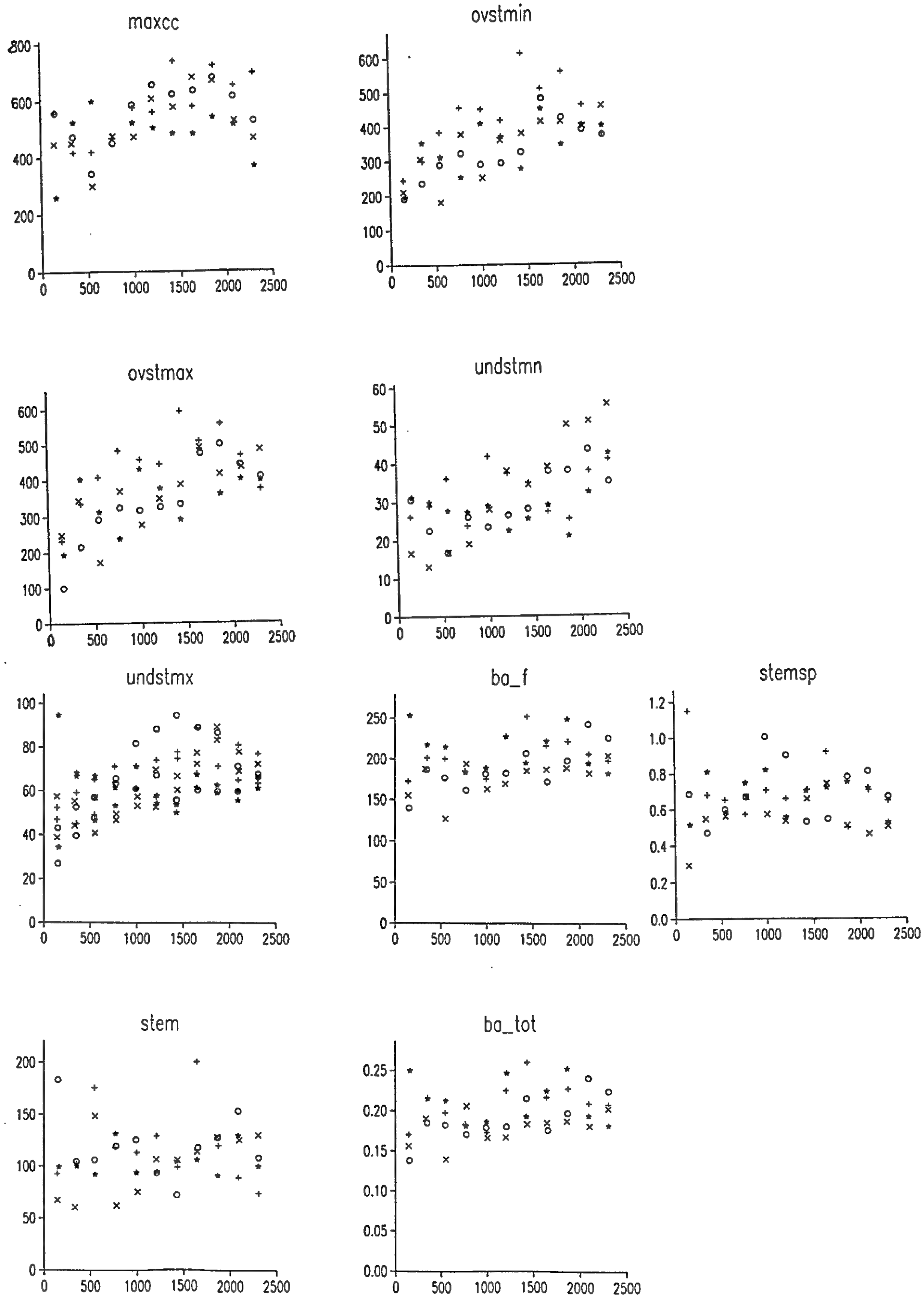


Figure 12: Directional experimental variograms for Set A variables for Surveys 1 and 2.

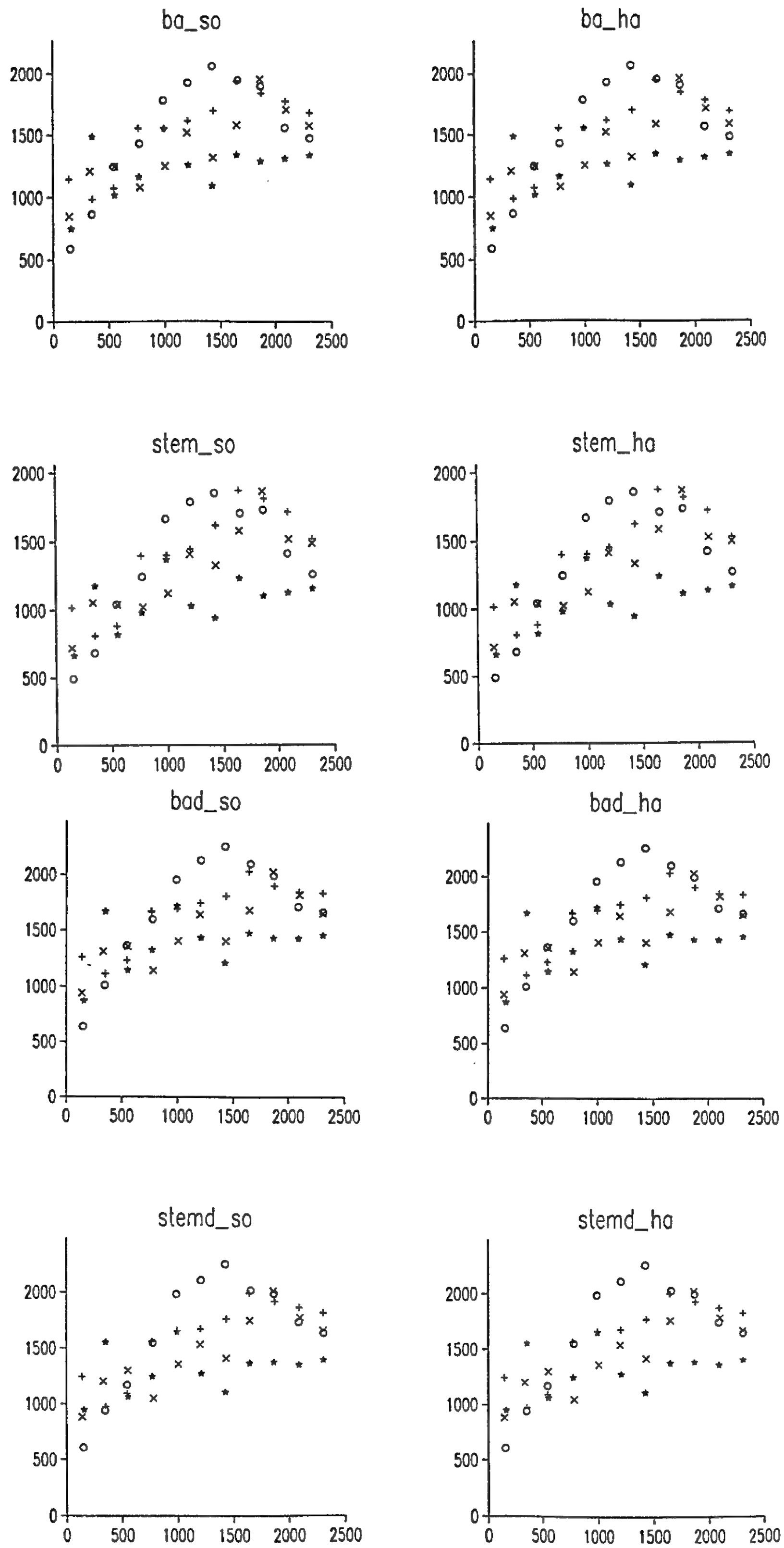


Figure 13: Directional experimental variograms for Set B variables for Surveys 1 and 2.

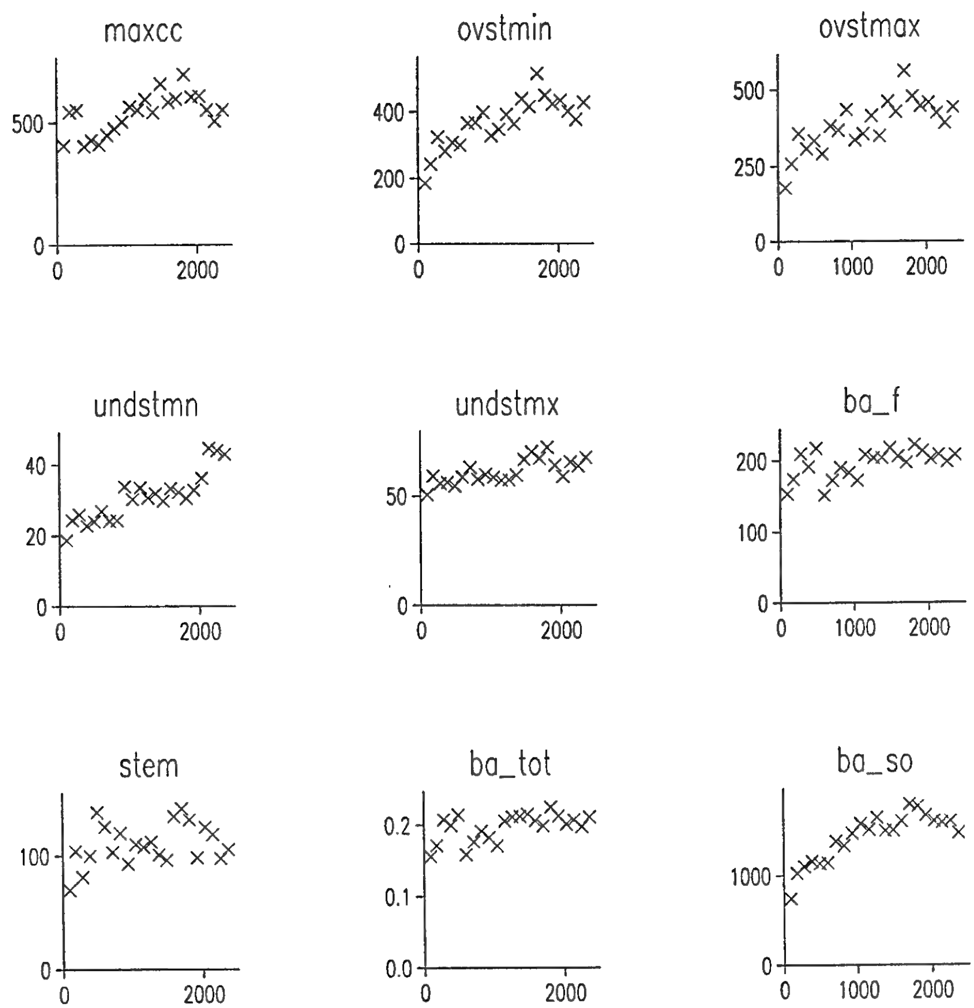


Figure 14: Omnidirectional experimental variograms of Set A variables from Surveys a 1 and 2.

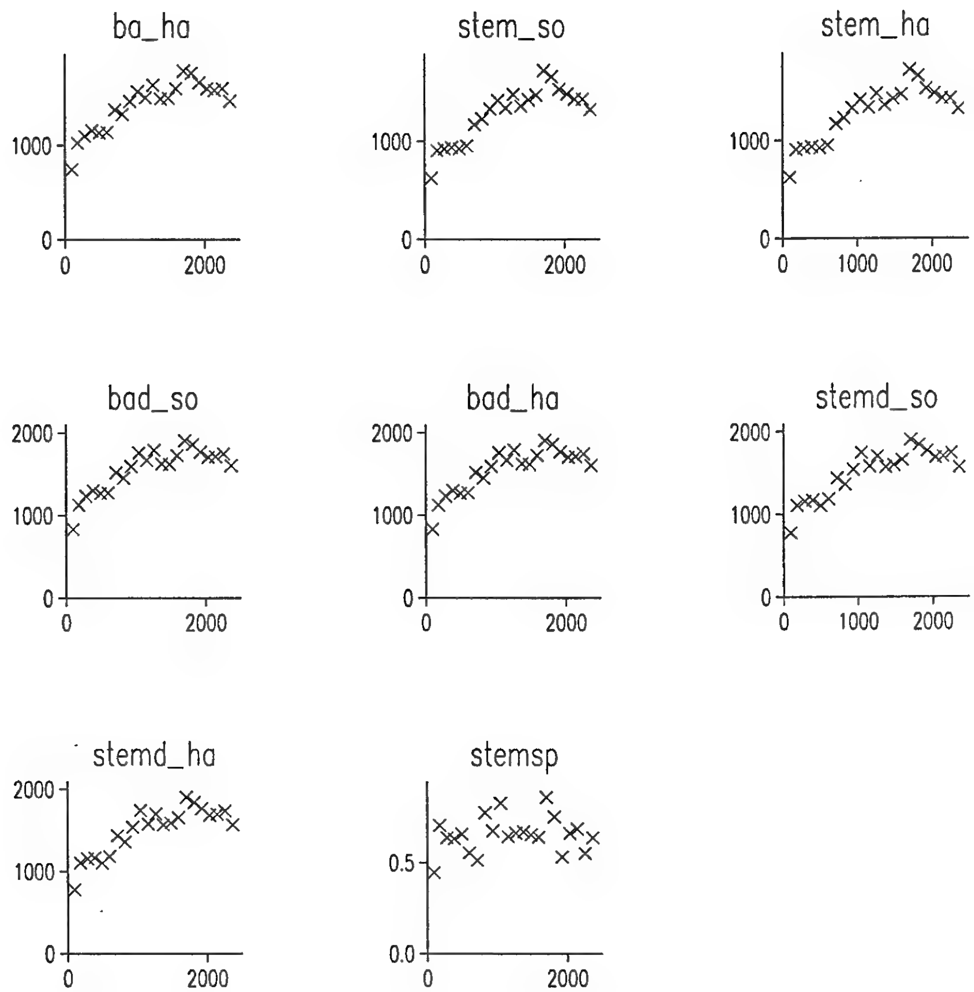


Figure 15: Omnidirectional experimental variograms of Set B variables from Surveys 1 and 2.

Table 9. Variogram model parameters.

Variables	*Model type*	*Nugget variance*	*Sill c_1*	*Sill c_2*	*Range a_1 (m)*	*Range a_2 (m)*
Canopy closure	Double spherical	0	397.4	218.3	129.0	1730.0
Overstory height	Circular	237.9	212.3		1850.0	
Understory height	Circular	19.2	12.6		1391.0	
Basal are (field)	Spherical	69.9	126.5		232.0	
Stem	Pentaspherical	21.4	65.7		380.0	
Stem spacing	Circular	0.305	0.193		407.0	
ba_so.ha	Double spherical	0	980.2	563.5	182.0	1553.0
ba_so/ha (45°)	Circular	662.3	1271.0		1271.0	
ba_so/ha (135°)	Circular	428.0	841.4		462.0	
stem_so/ha	Spherical	892.7	838.3		1428.0	
bad_so/ha	Spherical	909.1	819.9		1274.0	
stemd_so/ha	Circular	839.9	869.7		1432.0	
Multivariate variogram	Circular	6.05	3.08		309.9	
Elevation	Circular	93.05	312.5		1562.0	
Red (1)	Pentaspherical	2.21	17.89		906.0	
Green (2)	Double spherical	0	20.6	19.7	386.0	1047.0
NIR (3)	Circular	99.24	145.0		673.6	
NDVI	Double spherical	0.0015	0.00307	0.00202	666.8	1261.0

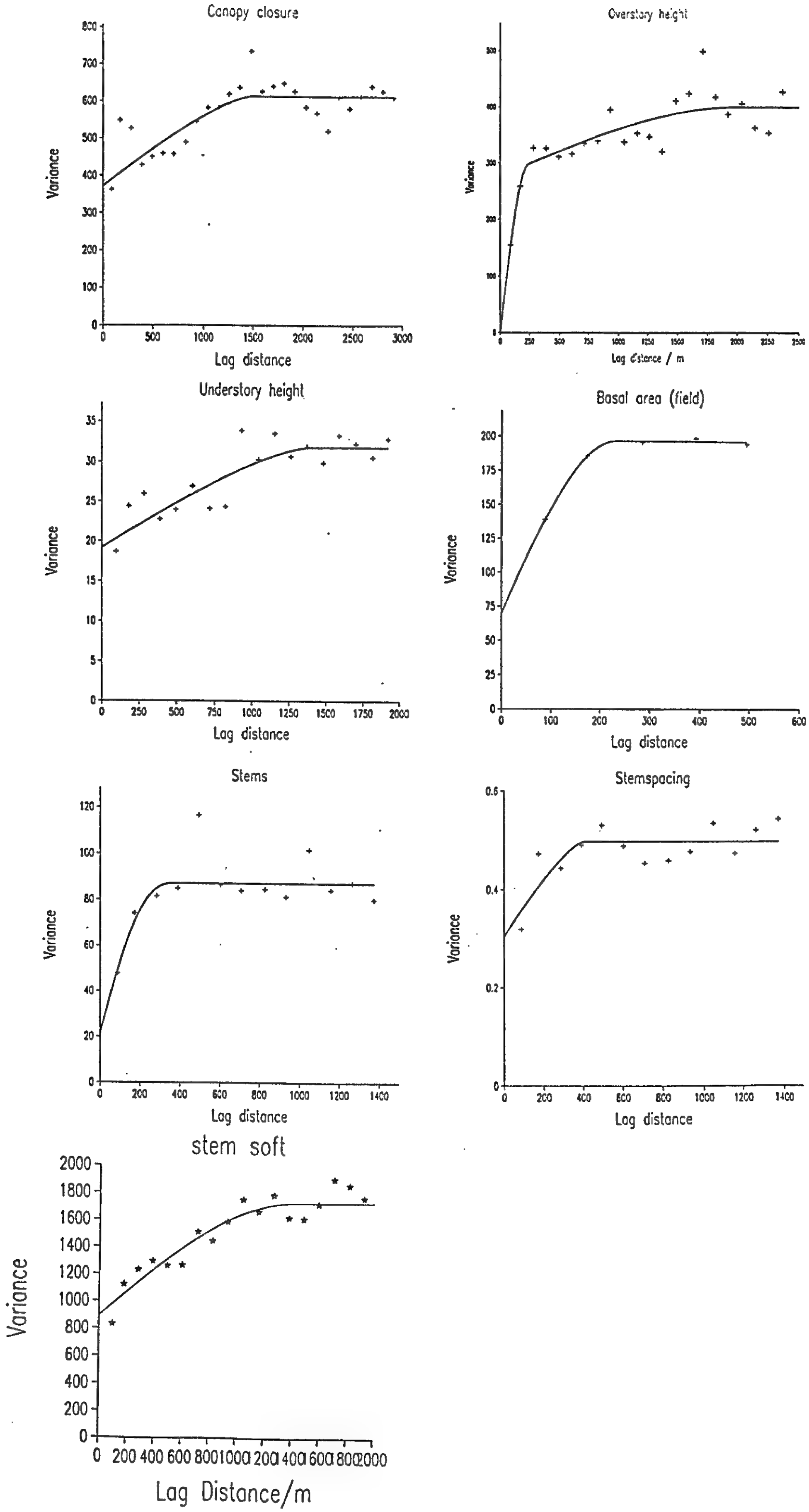


Figure 16: Experimental variograms and fitted models

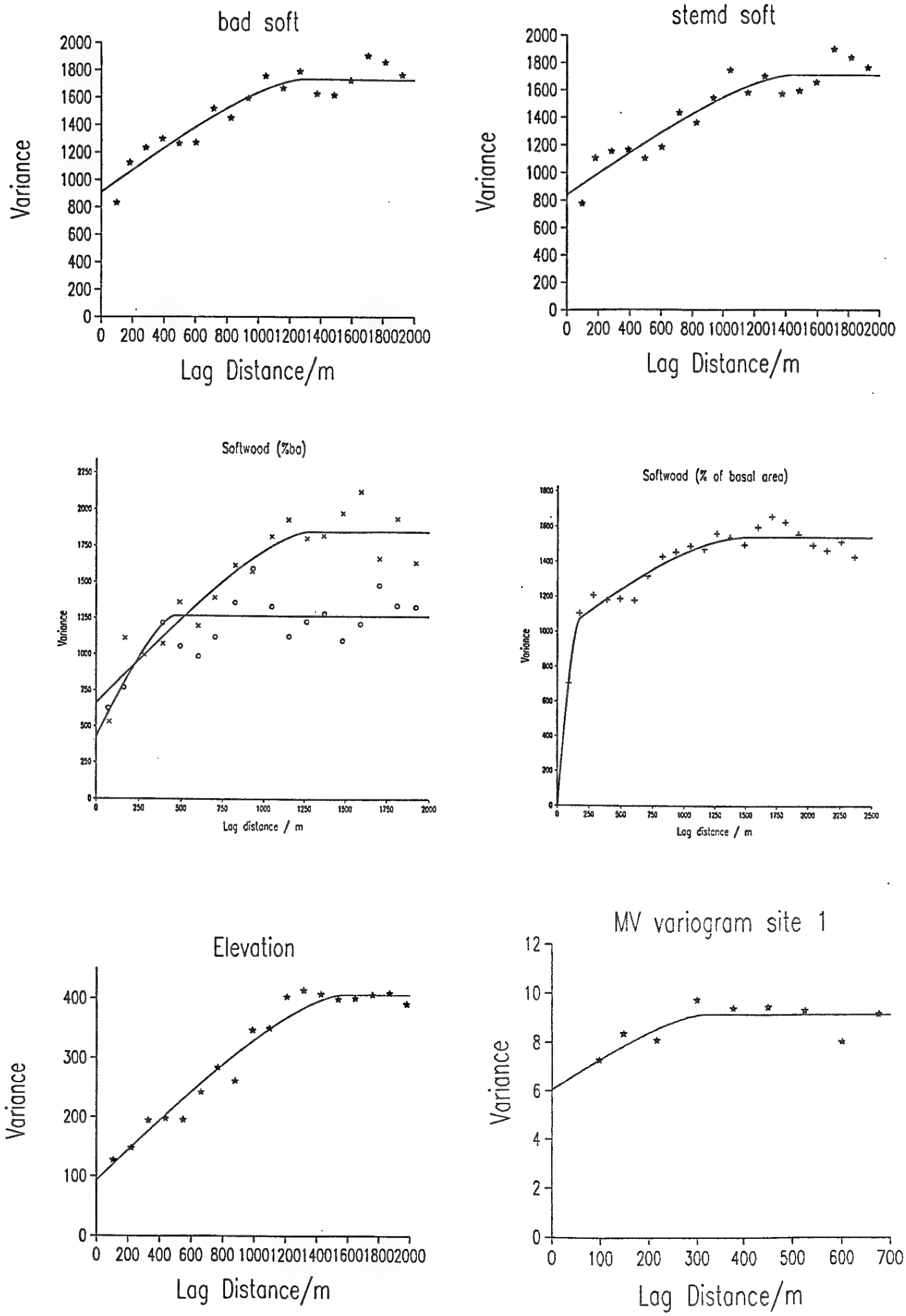


Figure 17: Experimental variograms and fitted model

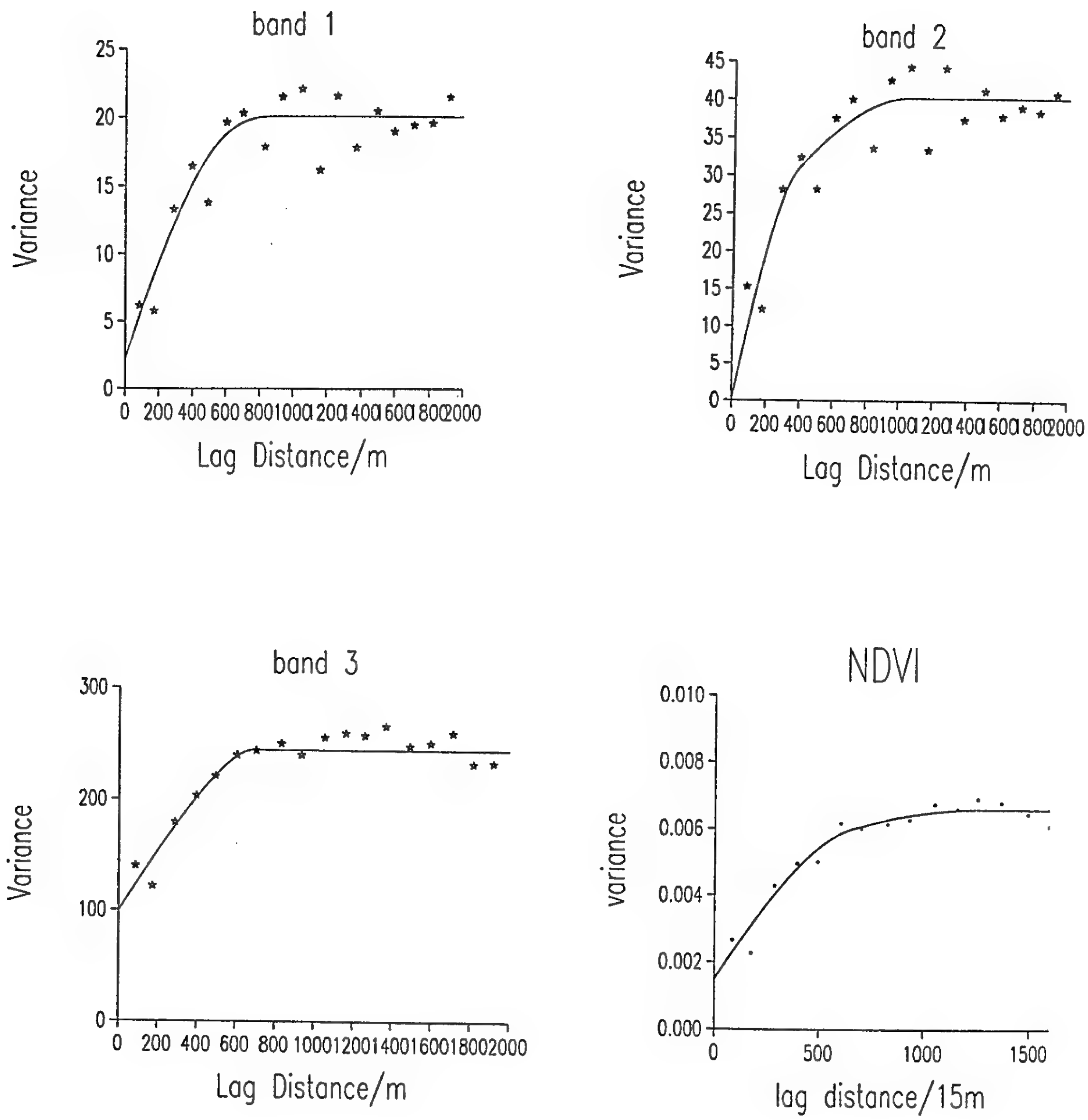


Figure 18: Experimental variograms and fitted models

Cross variograms

The theory for computing cross variograms between two or more variables is given at the beginning of the report. Cross variograms were computed between the vegetation measures and the values from the three SPOT wavebands. Those selected and shown in Figures 19 to 22 show some relation between the variables. For band 1 (Red) there is a negative relation between maxcc, unstmn and stem, and a positive relation between stem spacing (Figure 18). The relations with the other variables is not clear. For band 2 (Green) there are clear negative relations with maxcc and stem, and a positive relation with stem spacing (Figure 20). For band 3 (NIR) there are positive relations between maxcc, ovstmax, stem and ba_f, and a negative relation with ba_so (Figure 21). Cross variograms with elevation are give in Figure 22. Overall their relations with the vegetation measures are weak.

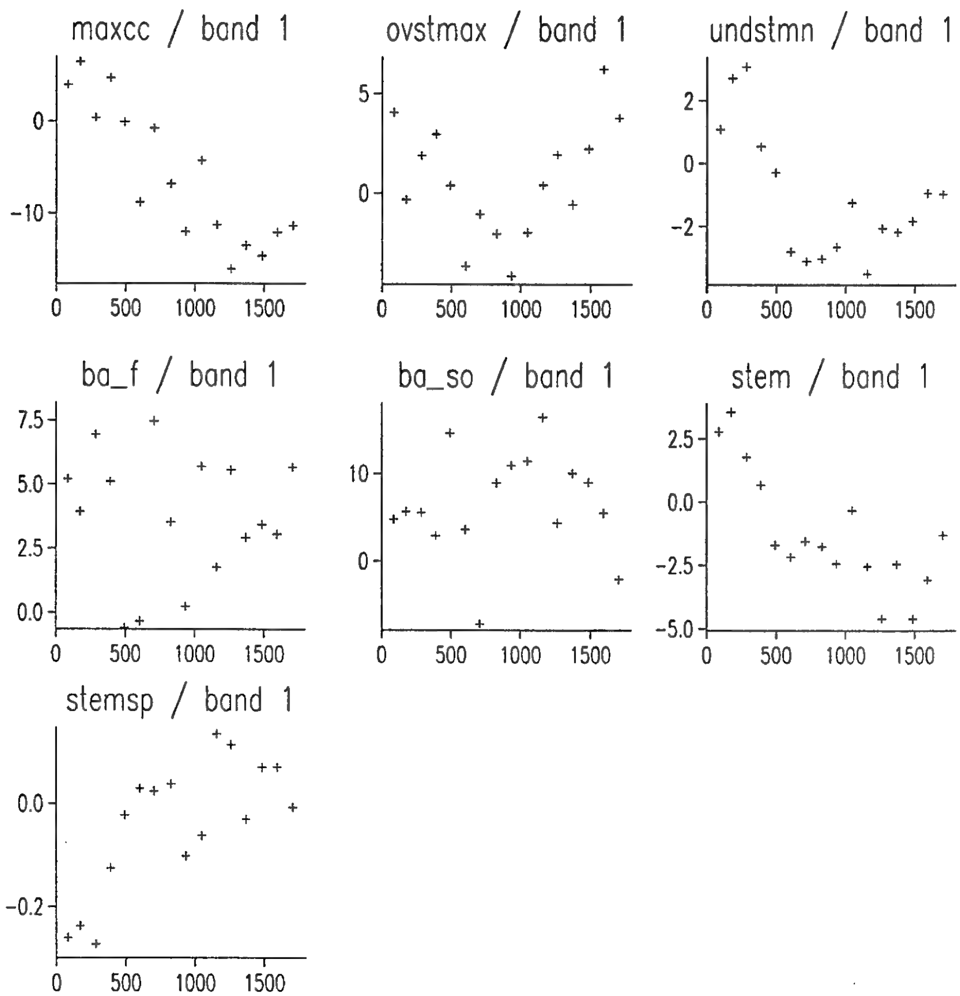


Figure 19: Cross experimental variograms between band 1 (Red) and selected vegetation measures.

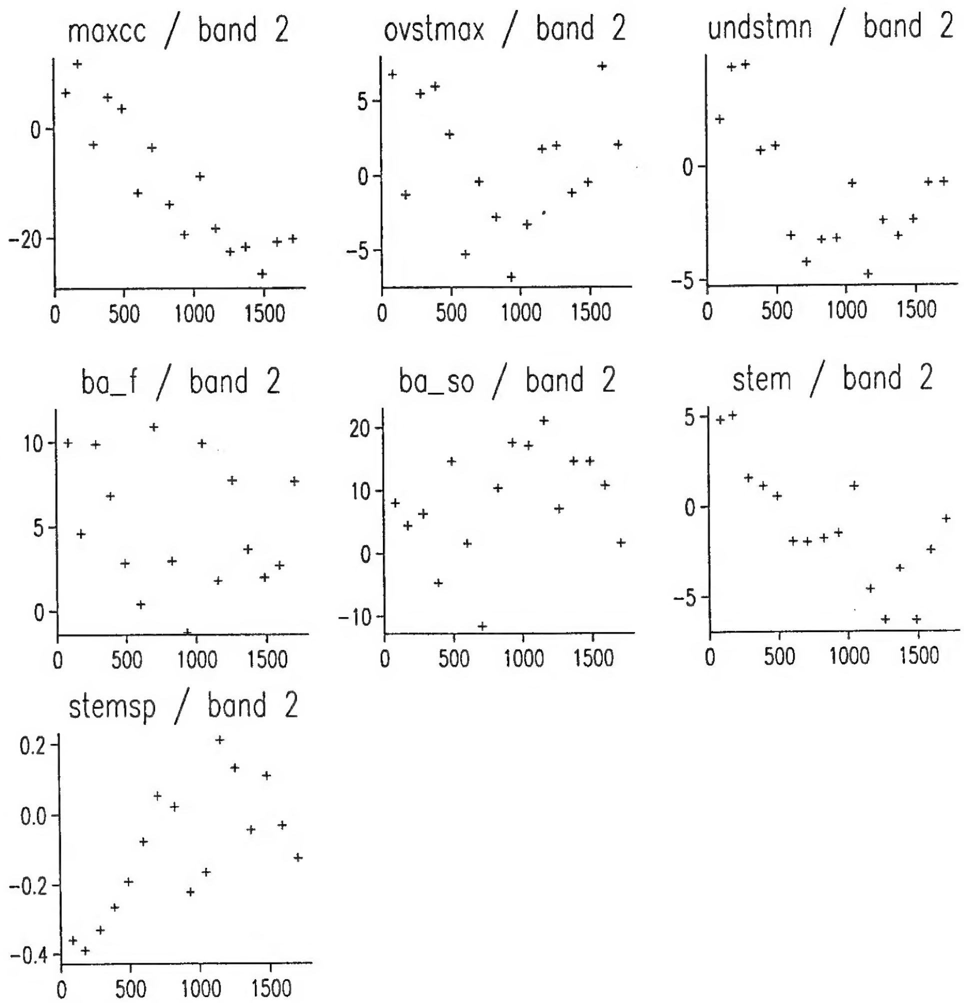


Figure 20: Cross experimental variograms between band 2 (Green) and selected vegetation measures.

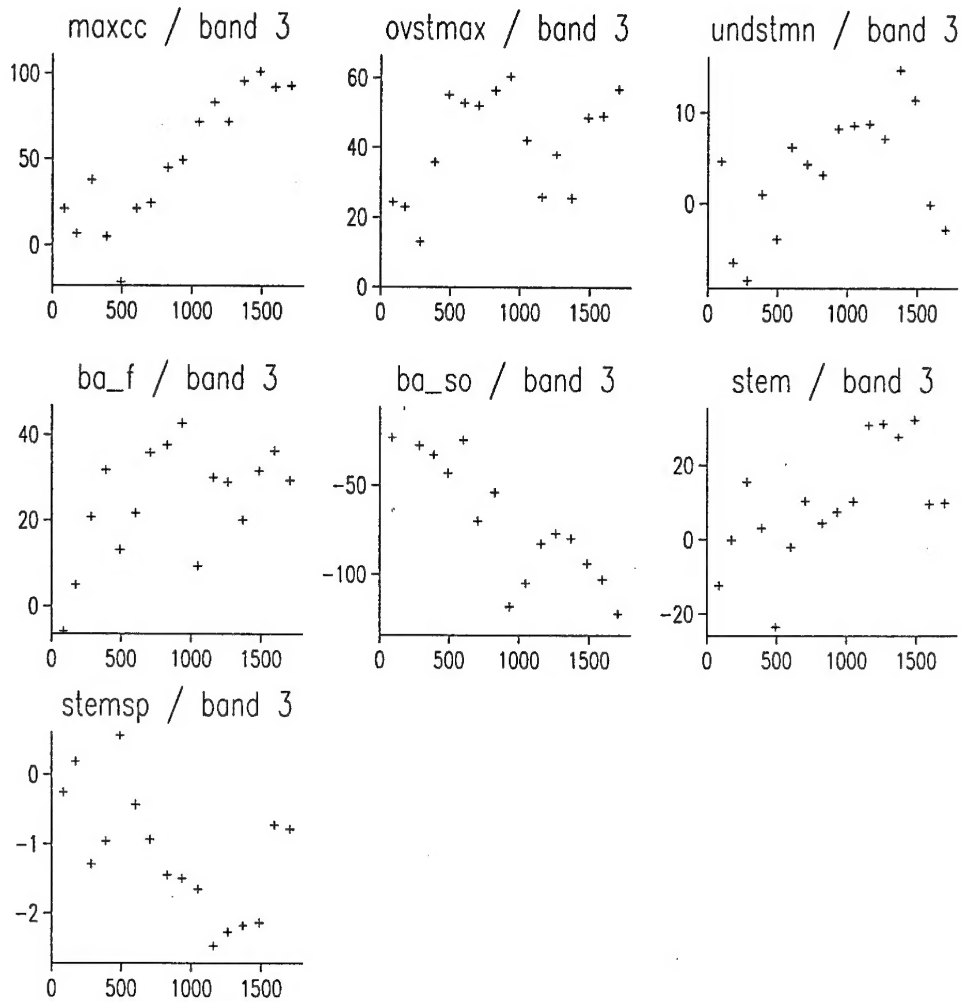


Figure 21: Cross experimental variograms between band 3 (NIR) and selected vegetation measures.

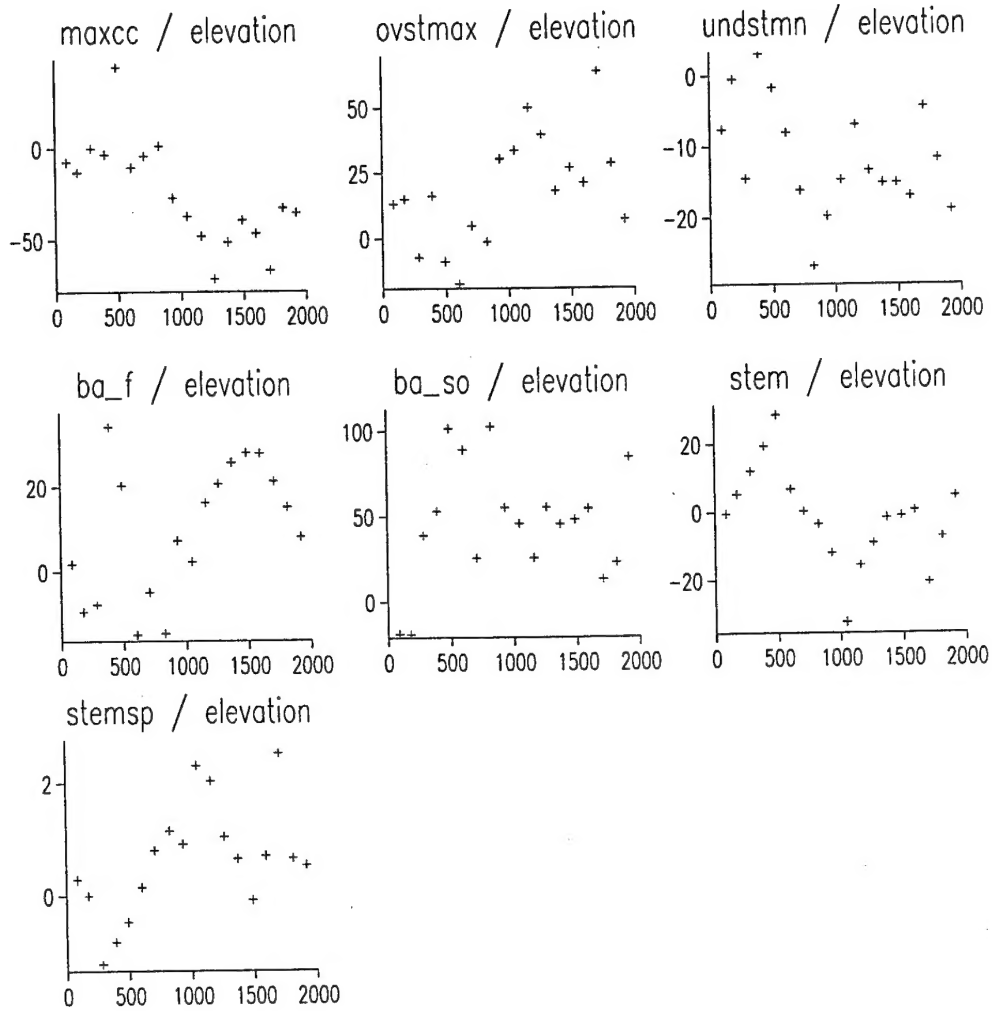


Figure 22: Cross experimental variograms between elevation and selected vegetation measures.

Acknowledgements

Johan Stendahl from the Agricultural University of Uppsala, Sweden, has analysed the data from the vegetation surveys and the image data that coincided with the sites. Ruth Kerry, soon to be a PhD student at the University of Reading, has helped to assemble the report and is currently analysing some of the data for the next report.